19th International Command and Control Research and Technology Symposium

"C2 Agility: Lessons Learned from Research and Operations"

Probabilistic Ontology Architecture for a Terrorist Identification Decision Support System

Topic 3: Data, Information,and Knowledge
Topic 5: Modeling and Simulation
Topic 2: Organizational Concepts and Approaches

Richard Haberlin
EMSolutions, Inc.
Arlington, Virginia
rjhaberlin@comcast.net

Paulo Cesar G da Costa
Kathryn B. Laskey
Systems Engineering and Operations Research
George Mason University
Fairfax, Virginia
[pcosta, klaskey]@gmu.edu

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **JUN 2014**	2. REPORT TYPE	3. DATES COVERED **00-00-2014 to 00-00-2014**
4. TITLE AND SUBTITLE **Probabilistic Ontology Architecture for a Terrorist Identification Decision Support System**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **EMSolutions, Inc,1401 South Clark Street Suite 200,Arlington,VA,22202**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

Presented at the 18th International Command & Control Research & Technology Symposium (ICCRTS) held 16-19 June, 2014 in Alexandria, VA. U.S. Government or Federal Rights License

14. ABSTRACT

Whether by nature or design, the personas of terrorists are often shrouded in mystery until they commit an act on the international stage. Without comment on the ethical dilemma that some identify with the practice, creating a profile of a terrorist from the available population serves as a starting point to reduce the volume of individuals requiring further investigation by limited analytic resources. A Terrorist Identification Probabilistic Ontology can assist the intelligence community in determining the likelihood of an individual being involved in terrorism using information about an individual?s relations, group associations, communications, and background influences. Intelligence analysts may use the proposed decision support system to identify those individuals that bear further scrutiny and pose a risk to target countries or their interests. Using the Reference Architecture for Probabilistic Ontology Development as a blueprint, an architecture is instantiated to develop a Terrorist Identification Probabilistic Ontology used for decision support. Ontologies are a fundamental enabling technology for system interoperability. They provide machine-interpretable representation of domain semantics, thus allowing interchange of information with unambiguous, shared meaning. However, a fundamental aspect of many real-world problems is uncertainty, which traditional ontologies do not represent. Representation of uncertainty in real-world problems requires probabilistic ontologies, which integrate the inferential reasoning power of probabilistic representations with the first-order expressivity of ontologies. The Reference Architecture for Probabilistic Ontology Development (RAPOD) catalogues and defines the processes and artifacts necessary for the development, implementation and evaluation of explicit, logical and defensible probabilistic ontologies developed for knowledge-sharing and reuse in a given domain. This paper provides an example implementation of the RAPOD in the form of an architecture for a Terrorist Identification Decision Support System.

15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT **Same as Report (SAR)**	18. NUMBER OF PAGES **28**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Probabilistic Ontology Architecture for a Terrorist Identification Decision Support System

Richard J. Haberlin, Jr.
EMSolutions, Inc.
Arlington, Virginia
rjhaberlin@comcast.net

Paulo C. G. da Costa
Kathryn B. Laskey
Systems Engineering and Operations Research
George Mason University
Fairfax, Virginia
[pcosta, klaskey]@gmu.edu

Abstract - **Whether by nature or design, the personas of terrorists are often shrouded in mystery until they commit an act on the international stage. Without comment on the ethical dilemma that some identify with the practice, creating a profile of a terrorist from the available population serves as a starting point to reduce the volume of individuals requiring further investigation by limited analytic resources. A Terrorist Identification Probabilistic Ontology can assist the intelligence community in determining the likelihood of an individual being involved in terrorism using information about an individual's relations, group associations, communications, and background influences. Intelligence analysts may use the proposed decision support system to identify those individuals that bear further scrutiny and pose a risk to target countries or their interests. Using the Reference Architecture for Probabilistic Ontology Development as a blueprint, an architecture is instantiated to develop a Terrorist Identification Probabilistic Ontology used for decision support. Ontologies are a fundamental enabling technology for system interoperability. They provide machine-interpretable representation of domain semantics, thus allowing interchange of information with unambiguous, shared meaning. However, a fundamental aspect of many real-world problems is uncertainty, which traditional ontologies do not represent. Representation of uncertainty in real-world problems requires probabilistic ontologies, which integrate the inferential reasoning power of probabilistic representations with the first-order expressivity of ontologies. The Reference Architecture for Probabilistic Ontology Development (RAPOD) catalogues and defines the processes and artifacts necessary for the development, implementation and evaluation of explicit, logical and defensible probabilistic ontologies developed for knowledge-sharing and reuse in a given domain. This paper provides an example implementation of the RAPOD in the form of an architecture for a Terrorist Identification Decision Support System.**

Keywords—probabilistic ontology, terrorism, inferential reasoning, architecture

I. INTRODUCTION

A. Background

Whether by nature or design, the personas of terrorists are often shrouded in mystery until they commit an act on the international stage. A decision support system (DSS) that draws upon existing reference knowledge coupled with current intelligence information can aid the intelligence analyst by reducing the number of individuals requiring further investigation by limited analytic resources. A Terrorist Identification Probabilistic Ontology (TIDPO) can provide a means to capture and catalog attributes and relationships of individuals, allowing probabilistic inference to identify suspects requiring further scrutiny. Specifically, the TIDPO can assist the intelligence community by determining the likelihood of an individual being involved in terrorism using information about this person's relations, group associations, communications, and background influences. This paper introduces an architecture for development of a TIDPO for use in a DSS.

Ontologies are a fundamental enabling technology for system interoperability. They provide machine-interpretable representation of domain semantics, thus allowing interchange of information with unambiguous, shared meaning. However, a fundamental aspect of many real-world problems is uncertainty, which traditional ontologies do not represent. Therefore, a means to incorporate uncertainty is a necessity. Representation of uncertainty in real-world problems requires probabilistic ontologies (PO), which integrate the inferential reasoning power of probabilistic representations with the first-order expressivity of ontologies. Probabilistic ontologies extend current ontology formalisms to provide support for representing and reasoning with uncertainty.

Using the Reference Architecture for Probabilistic Ontology Development (RAPOD) [1] as a blueprint, the TIDPO Architecture is instantiated to develop a probabilistic ontology used for decision support. The architecture provides synergy of effort by identifying concepts, processes, languages, and tools for designing and maintaining the TIDPO. It details each of the components and defines the criteria to be satisfied by the selected tools and methods. Further, this architecture may be used to develop similar probabilistic ontologies for decision support in similar domains.

B. Scope

Terrorists maintain an unremarkable profile and utilize advanced social networking communications techniques to minimize the likelihood that they are detected before executing planned attacks. Extending the model introduced in [2], the PO conceptualized in this paper is applicable to terrorists that target the West as identified by Sageman [3]. Terrorists are commonly identified as multinational and transient, compounding factors in identifying potential terrorists from the multitude of persons that interact with U.S. interests, at home and abroad. However, using information about an individual's relations, group associations, communications, and background influences may provide insight into the likelihood of a person being involved in terrorism. While some affiliations may increase the likelihood that an individual may join a terrorist group and attempt access to a target country, there is always the uncertainty that comes from the human condition. Further, each individual may participate in multiple organizations (some of which may be associated with terrorism) or have multiple friends and relatives (some of whom may participate in terrorism). Uncertainty associated with the multitude of factors affecting the crewmember's context must be captured conditionally. Without comment on the ethical dilemma that some identify with the practice, creating a profile of a terrorist from available population data serves as a starting point to reduce the volume of individuals requiring further investigation by limited analytic resources. The TIDPO will incorporate domain knowledge and individual attributes to infer the likelihood an individual is involved in terrorism and therefore bears further scrutiny.

A decision support system is an interactive, computer-based information system that supports business or organizational decision-making activities through compilation, processing and display of domain information. Its purpose is to assist in the activity of decision making by providing an organized set of tools intended to impose structure on portions of the situation and to improve the ultimate effectiveness of the decision outcome [4]. With the ever-increasing volume of information delivered to the analyst, there is a need for advanced decision support through data compilation, screening, transformation and probabilistic inference. Input may include raw data, documents, interviews, and mathematical models stored in databases, ontologies, and probabilistic ontologies. Because each DSS is domain-specific, it has a narrow focus of applicability and will only address a narrow set of decisions. In this paper, the DSS for terrorist identification is the desired product for the intelligence analyst, supported through implementation of the TIDPO specified in the architecture. The TIDPO will be populated using the work of Marc Sageman [3] to validate the model. Incorporating biographical data for the 172 terrorists studied, subsequent work will test the PO model against the 911 terrorists to see if the TIDPO correctly identifies the perpetrators as needing further scrutiny.

C. *Model Implementation and Viewpoint*

The concept model shown in Figure 1 illustrates the scope of the TIDPO Architecture supporting the Terrorist Identification DSS in which an intelligence analyst is aided in producing a contextually driven decision. The DSS is updated using available data regarding the current operational environment and intelligence. It is based on a knowledge base and supported by a PO grounded in the reference environment. The architecture described below is a blueprint for development of the TIDPO to support the DSS.

As shown in the figure, an ontology of relevant, hierarchical relationships among terrorism-associated classes is constructed. Then, uncertainty is introduced based on a reference environment representing a contextually relevant situation. For example, the intelligence analyst may be interested in terrorism based in the Arabian Peninsula. This would guide queries to the Intelligence Knowledge Base for relevant relationships. Evidence from the available knowledge base is applied to the probabilistic ontology to provide the DSS with inferential reasoning support that is tailored for the chosen operational domain. After implementation and during operations, the DSS continually receives updated information about the current operational situation and changes to the environment. These data update the intelligence knowledge base, and therefore the probabilistic ontology. The end result is a DSS that produces contextually-driven decisions about the domain of interest using both historical and current evidence.

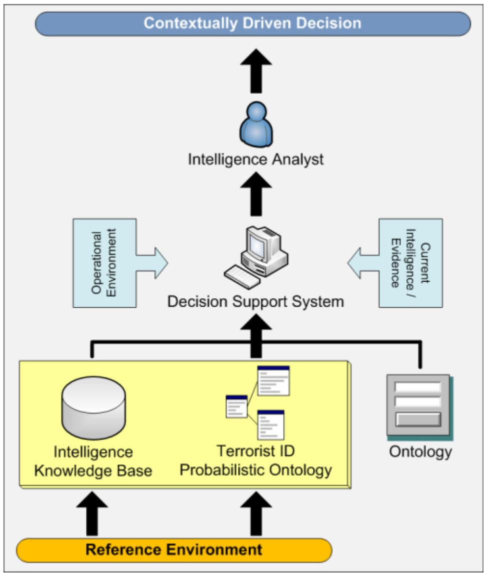

Contextually Driven Decision
Intelligence Analyst
Operational Environment
Decision Support System
Current Intelligence / Evidence
Intelligence Knowledge Base
Terrorist ID Probabilistic Ontology
Ontology
Reference Environment

Figure 1 - Concept Diagram for Terrorist Identification DSS

At the highest level of abstraction, the TIDPO architecture responds to a requirement for decision support by the Intelligence Analyst. Specifically, it describes the composition of the system by providing determination of structural elements, their interfaces, and their behavior [5]. The architecture codifies captured lessons learned and best practices, acknowledges wisdom and presents a set of services, design concepts, components and configurations applicable to the specific domain of interest. Creating an architecture for a given domain problem results in a reusable blueprint for similar designs that facilitates successful development from conceptualization to operation. Using the RAPOD, an architecture is instantiated for the TIDPO, illustrated in Figure 2 .

II. Probabilistic Ontology Architecture for a Terrorist Identification Decision Support System

The PO architecture in Figure 2 illustrates the TIDPO from conceptualization as a DSS that is required to determine likelihood of terrorist affiliation to the operational implementation of a PO that performs inferential reasoning to support that requirement. In the Input Layer, references to appropriate tables detailed below lead to specification of objectives, requirements, metrics and rules. Similarly, in the Methodology Layer ontology reuse, the Probabilistic Ontology Development Methodology (PODM) [6], ontological engineering, and learning are linked with their descriptions in Section II.B. Neither ontological learning nor probabilistic

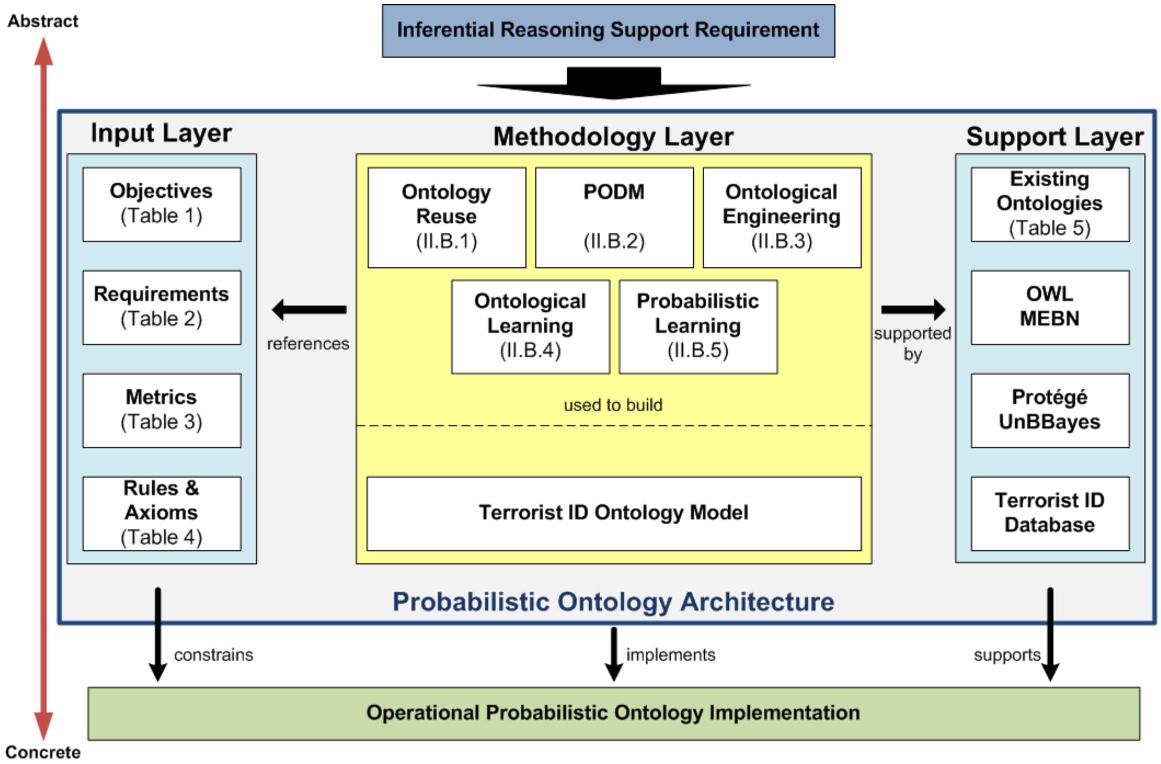


Figure 2 – Terrorist Identification Probabilistic Ontology Architecture

learning is used for this instantiation. A database of 172 known terrorists was constructed as the knowledge base, which captures multiple attributes for each individual. Ontological engineering is performed on this KB to create the Terrorist Identification Ontology in Protégé. The Support Layer consists of technological artifacts highlighted by the OWL and MEBN languages used to represent the ontology and probabilistic ontology, respectively. Software included Protégé for ontology modeling and UnBBayes for probabilistic ontology modeling. Finally, the research of Marc Sageman is used to generate the Terrorist Database. From this blueprint, the TIDPO will be developed using the PODM specified in [6].The TIDPO is created by ingesting Terrorist Identification Ontology and incorporating uncertainty in MEBN using UnBBayes. A probabilistic ontology provides a means to represent and reason with uncertainty by integrating the inferential reasoning power of probabilistic languages with the first-order expressivity of ontologies. Few things are certain, and inferring in the presence of uncertainty allows the analyst to focus attention on the most relevant data through designed queries.

A. Input Layer

The Input Layer defines external influences on the probabilistic ontology and is referenced by components of the Methodology Layer. It contains those components expected to provide detail on the purpose of the PO and its bounding constraints in the form of system requirements. Population of the Input Layer occurs primarily during the early stages of the development process during which the Stakeholder and Developer work closely to identify the objective of the model, expectations of its performance, and resource restrictions. Parameters specified in the Input Layer will constrain the operational implementation.

1) Objective. With an Intelligence Analyst stakeholder, the objective is given in Table 1. The domain for this instantiation of the TIDPO is Islamic fanatics of the global Salafi Jihad identified by Sageman in *[3]*.

Table 1 - Objective Statement

Objective
Provide decision support through inferential reasoning to determine the likelihood a particular individual is a terrorist based on a profile of background, relationships, communications, and associations.

2) Requirements. Requirements define the system to be implemented in terms of its behaviors, applications, constraints, properties, and attributes. Table 2 records the initial requirements elicited from the Stakeholder through an iterative process that included objective setting, background knowledge acquisition, knowledge organization, and requirements collection *[7]*.

Table 2 - Table of Selected Requirements

ID	Requirement
R1	**Determine likelihood individual is a terrorist**
R2	**Background**
R2.1	Ingest knowledge of killed in OEF
R2.2	Ingest knowledge of imprisoned in OEF
R2.3	Ingest family status
R2.4	Ingest place of worship
R2.5	Ingest former military/police
R2.6	Ingest government
R3	**Relationships**
R3.1	Ingest family involvement
R3.2	Ingest friend involvement
R3.3	Ingest social network information
R4	**Associations**
R4.1	Ingest nationality
R4.2	Ingest economic standing
R4.3	Ingest education level
R4.4	Ingest occupation
R5	**Communications**
R5.1	Ingest cell phone use
R5.2	Ingest email use
R5.3	Ingest weblog use
R5.4	Ingest chat room use
R6	**Performance**
R6.1	Must run on PC computer
R6.2	Must provide solution in 2 minutes

The goal of this task is to capture attributes that should be controlled within the model in written requirement statements, to be validated by the Stakeholder and measured by the metrics. The operational PO will be evaluated against these requirements.

3) *Metrics*. Metrics characterize the criteria by which the fielded system is to be evaluated. For the TIDPO, the primary metric of interest is *P(terrorist|background, relationships, associations, communications)*, which defines model accuracy. An initial set of metrics based on the requirements is captured in Table 3. It is best if there is at least one metric to support each requirement of the system.

4) *Rules and Axioms*. Formal Axioms are first-order logical expressions that are always true. Rules are used to infer attribute values, or relation instances *[8]*. The Formal Axioms and Rules Table also captures heuristics and algorithms that act as constraints for the model. Table 4 summarizes selected Axioms and Rules from the TIDPO.

These heuristics and algorithms are used as bounding constraints to scope the model appropriately for the domain by capturing plain-language relationship statements in machine-readable format. Relevant heuristics and algorithms are regarded as Axioms which are propositions assumed without proof for the sake of studying the consequences that follow from it [9].

B. *Methodology Layer*

The Methodology Layer contains the heart of the probabilistic ontology development process including the Probabilistic Ontology Development Methodology that allows creation of a specific probabilistic ontology implementation to support the requirements of the Stakeholder. The Methodology Layer references information gathered in the Input Layer and is assembled using components and tools from the Support Layer. Its individual components are introduced below.

1) *Ontology Reuse*. Before beginning construction of the ontology, it is useful to research existing ontologies in related domains to be reused and/or extended for the current problem. Model reuse is defined as the process by which available knowledge is used as input to generate new models. Reusing existing models may also require ontological re-engineering as described by Gomez-Perez et al. [8]. For the Terrorist Identification Ontology, three existing ontologies are reused by incorporating applicable classes and relations.

2) *Probabilistic Ontology Development Methodology*. Extending the work of Carvalho [27], the PODM completes the evolution of requirements into an ontology that is probabilistically-integrated. A probabilistically-integrated ontology combines the inferential reasoning power of probabilistic representations with the first-order expressivity of ontologies. A key component of that methodology is a detailed Construction Process, which explicitly describes the iterative tasks required to produce a probabilistic ontology with in-situ evaluation steps to ensure continuous operation for inferential reasoning.The PODM will be used to perform the iterative construction that extends the Terrorist Ontology to incorporate uncertainty, creating the TIDPO.

3) *Ontological Engineering*. An ontology is used to capture consensual knowledge about a domain of interest [8]. Selection of the appropriate ontological engineering methodology is context dependent as is the required fidelity of the ontological model. Terms and processes for development are as various as the application for which they are used. A generalized sequence of steps iteratively modeled for ontological engineering is proposed below in Figure 3.

Table 3 - Table of Selected Metrics

Requirement		Metric			
ID	Name	ID	Name	Definition	Units
R1	Terrorist Individual	M1	Model Accuracy	Correctly identify the likelihood (≥ 85%)	Percent
R2	Background	M2	Model Flexibility	Ingest/operate on ontology of 172 individuals	Items
R3	Relationships				
R4	Associations				
R5	Communications				
R6	Performance	M3	Execution Time	Generate solution in 2 minutes or less	Min
R6	Performance	M4	Model Efficiency	Compute solution on pc computer (Intel 1.3GHz)	Processor

Table 4 - Formal Axioms and Rules

Axiom	Nationality	Names	Communication	Terrorist
Description	Each individual is associated with a single nation	Each individual is known by a single name	A terrorist will communicate with certainty	There is a possibility that any individual in demographic is a terrorist
Expression	NA	NA	P(communicate) = 1.0	P(Terrorist) = 0.001
Classes	Person Nation	Person	Person	Person
Relations	hasNationality	hasName	NA	NA
Variables	NA	NA	ComWithTerrorist	isTerrorist

Ontological Engineering Process

i. **Identify Classes:** *what objects are acting or acted upon?*

i. **Develop Context:** *where or when are the actions occurring?*

i. **Identify Relationships:** *what objects are affected by an object?*

/. **Identify States:** *in what condition may an object be found?*

Figure 3 - Ontological Engineering Process

Ontological engineering ensures the development of an explicit, logical and defensible ontologies for knowledge-sharing and reuse that will be extended to become the TIDPO.

4) *Ontological Learning*. There are several methods to aid in the knowledge acquisition process required to build an ontology. Ontological Learning was not employed in the TIDPO instantiation.

5) *Probabilistic Learning*. For the TIDPO model, local probability distribution (LPD) values are given by the domain research conducted by Sageman. However, probabilistic learning would be a means to incorporate data from additional individuals for an extended knowledge base.

6) *Ontology*. The Terrorist Identification Ontology is created in OWL using Protégé. The working ontology serves as the relational framework for the PO when uncertainty is introduced. Construction tools and environments such as Protégé [10] aid in the key ontological engineering tasks of implementation, consistency checking, and documentation. At this point the ontology is implemented in a suitable ontology building environment and evaluated for consistency. For this project, the Protégé (Version 4.1) ontology development environment is used to capture terrorist identification domain information [10].

C. *Support Layer*

The Support Layer provides the background technology and design strategy necessary to instantiate the conceptualization of a specific probabilistic ontology to satisfy identified requirements. It includes existing ontologies available for reuse or re-engineering, software tools that enable ontology and probabilistic ontology development, mathematical languages that allow representation of entity attributes and their relationships, and databases of existing facts referenced for learning and knowledge base population. The purpose of the Support Layer is to facilitate probabilistic ontology development by identifying technological and semantic features specific to a particular inferential reasoning model. The four Support Layer components are discussed below.

1) *Existing Ontologies*. Existing ontologies were available for reuse as shown in Table 5.

Table 5 - Existing Ontologies

Ontology	Utility
geopolitical.owl	Nations, groups, neighbors
Generations.owl	Family relationships
Biography.owl	Individual personal data

As previously discussed in Section II.B.1, model reuse is a strength of the ontological engineering discipline and effort should be made to research and incorporate existing ontology material into new application areas.

2) Modeling Languages. Ontological engineering was conducted in the Web Ontology Language (OWL) due to its incorporation within Protégé and UnBBayes software tools. All of the existing ontologies used were modeled in OWL. Multi-Entity Bayesian Networks (MEBN) was used for probabilistic ontology development due to the maturity of available software tools, specifically UnBBayes.

3) Software Tools. While there are several software tools available for ontological engineering, at this time only UnBBayes is mature enough to produce working probabilistic ontologies. UnBBayes ingests an OWL ontology and extends it to account for uncertainty. Therefore, Protégé is used to capture the OWL ontology and UnBBayes for the probabilistic ontology.

4) Knowledge Base. The knowledge captured in the Terrorist Ontology is primarily gleaned from the work of Sageman [3]. It includes data about 172 terrorists and includes information about their geographical origins, socioeconomic status,education, faith, occupation, family status, psychology, age, employment, friendship, kinship, discipleship, social network, etc.

III. CONCLUSION

A. *Summary*

Since the terrorist attacks of September 11, 2001, there has been a great deal of interest in expeditious determination of the composition, operations and resourcing of terrorist networks. In the information technology domain, much of the focus has been on mining open-source material such as email, weblogs, and news articles to build a representation of terrorist social, resource, and operational networks. A Decision Support System that combines information about relations, group affiliations, communications, and ethno-religious or political background into a model describing the likelihood that a particular individual becomes a terrorist will provide the intelligence analyst with a powerful tool to prioritize limited investigative resources. The architectural introduced in this paper provides a blueprint to develop the probabilistic ontology needed to support this tool through inferential reasoning.

B. *Future Work*

Continuation of this work will include instantiation of the probabilistic ontology and eventual testing against the personal profiles of the known 9/11 terrorists. Using the classes, relationships, and probabilities identified by Sageman [3], the terrorist ontology will be instantiated and uncertainty applied by extending the model introduced in [2] and following the Probabilistic Ontology Development Methodology [6]. This working probabilistic ontology will be evaluated by instantiating evidence statements for each of the 19 terrorists associated with the attack on September 11, 2001.

IV. REFERENCES

[1] Richard J. Haberlin, Paulo C.G. Costa, and Kathryn B. Laskey, "A Reference Architecture for Probabilistic Ontology Development," in *Proceedings of the 8th International Conference on Semantic Technology for Intelligence, Defense, and Security*, Fairfax, VA, 2013, pp. 1-9.

[2] Richard J. Haberlin and Paulo C.G. Costa, "A Bayesian Model for Determining Crew Affiliation with Terrorist Organizations," in *Proceedings of the Quantitative Methods in Defense and National Security*, Fairfax, 2010, p. 10.

[3] Marc Sageman, *Understanding Terror Networks.* Philadelphia: University of Pennsylvania Press, 2004.

[4] George M. Marakas, Decision Support Systems in the 21st Century. Upper Saddle River: Prentice Hall, 2003.

[5] Philippe Kruchten, The Rational Unified Process: An Introduction. Upper Saddle River: Addison-Wesley, 2004.

[6] Richard J. Haberlin, Probabilistic Ontology Reference Architecture and Design Methodology, 2013, PhD Dissertation.

[7] Ian Kotonya and Ian Sommerville, Requirements Engineering. Chichester: John Wiley & Sons, 1998.

[8] Asuncion Gomez-Perez, Fernandez-Lopez Mariano, and Oscar Corcho, Ontological Engineering with Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web. London: Springer-Verlag, 2010.

[9] Dictionary.com, LLC. (2013, June) "axiom" in Dictionary.com Unabridged. [Online]. http://dictionary.reference.com/browse/axiom?s=t

[10] Stanford University. (2011, January) Protege. [Online]. http://protege.stanford.edu/

[11] Ahmed E. Hassan and Richard C. Holt, "A Reference Architecture for Web Servers," in Proceedings of the Seventh Working Conference on Reverse Engineering, Brisbane, 2000.

[12] Dictionary.com LLC. (2013) Dictionary.com. [Online]. http://dictionary.reference.com/

[13] Thomas R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," International Journal of Human-Computer Studies, pp. 907-928, 1995.

[14] Grigoris Antoniou and Frank Van Harmelen, "Web Ontology Language: OWL," in Handbook on Ontologies in Information Systems.: Springer-Verlag, 2003.

[15] Paulo Cesar G. da Costa. (2005, July) PhD George Mason Univeristy. [Online]. http://hdl.handle.net/1920/455

[16] IEEE, IEEE Standard Glossary of Software Engineering Terminology. New York: IEEE Computer Society, 1990.

[17] IEEE, IEEE Standard for Developing Software Life Cycle Processes. New York: IEEE Computer Society, 1996.

[18] Office of the Assistance Secretary of Defense for Networks and Information Integration (OASD/NII), "Reference Architecture Description," Arlington, 2010.

[19] William B. Rouse and Andrew P. Sage, Handbook of systems engineering and management. Hoboken: John Wiley & Sons, 2009.

[20] F. G. Patterson, "Systems Engineering Life Cycles: Life Cycles for Research, Development, Test, and Evaluation; Acquisition; and Planning and Marketing," in Handbook of Systems Engineering and Management.: John Wiley & Sons, 2009, pp. 65-115.

[21] A. Schauerhuber, W. Schwinger, E. Kapsammer, W. Retschitzegger, and M. Wimmer, "Towards a Common Reference Architecture for Aspect-Oriented Modeling," in Proceedings of the 8th International Workshop on Aspect-Oriented Modeling, Bonn, 2006.

[22] Heather Kreger, Vince Brunssen, Robert Sawyer, Ali Arsanjani, and Rob High. (2012, Jan) IBM Developer Works. [Online]. http://www.ibm.com/developerworks/webservices/library/ws-soa-ref-arch/

[23] Johan Eltes, The Reference Architecture - a foundation for successful projects, 2004.

[24] Alexander Levis, "System Architectures," in Handbook of Systems Engineering and Management.: John Wiley & Sons, 2009, pp. 479-506.

[25] Jr., James E. Armstrong, "Issue Formulation," in Handbook of Systems Engineering and Management. Hoboken: John Wiley & Sons, 2009, pp. 1027-1089.

[26] Alan Grosskurth and Michael W. Godfrey, "A Reference Architecture for Web Browsers," in International Conference on Software Maintenance, Budapest, 2005.

[27] Rommel Novaes Carvalho. (2011, June) PhD George Mason Univeristy. [Online].

BIOGRAPHIES

Richard J. Haberlin, Jr. is a Senior Operations Research Analyst and Subject-Matter Expert for EMSolutions, Inc. in Arlington, Virginia. Dr. Haberlin is a retired U.S. Naval Flight Officer with extensive experience in anti-submarine warfare and airborne intelligence, surveillance and reconnaissance operations in the Arctic, Atlantic, Mediterranean and Middle East. His research interests include inferential reasoning, probabilistic ontology development, Bayesian networks, and model-based systems engineering.

Kathryn Blackmond Laskey is Professor of Systems Engineering and Operations Research and Associate Director of the Center of Excellence in Command, Control, Communications, Computing and Intelligence at George Mason University. Dr. Laskey teaches courses in systems engineering, computational decision theory, and decision support. She has published extensively in the areas of inference, knowledge representation, learning, and information fusion. She developed Multi-Entity Bayesian Networks (MEBN), a language and logic that extends first-order logic to support probability. She was a key contributor to the development of PR-OWL, an upper ontology that allows MEBN theories to be represented in OWL ontologies.

Paulo Cesar G Costa is Associate Professor of Systems Engineering and Operations Research and Research Director for C2 Activities of the Center of Excellence in Command, Control, Communications, Computing and Intelligence at George Mason University. Dr. Costa is a retired Brazilian Air Force Flight Officer with extensive experience in electronic warfare, C4I, operations research and military decision support. He teaches courses in decision theory and systems engineering, and has developed PR-OWL, a probabilistic extension of the OWL ontology language. As an invited professor at University of Brasilia, he was a key contributor to the development of UnBBayes-MEBN, an implementation of the MEBN probabilistic first-order logic.

Probabilistic Ontology Architecture for a Terrorist Identification Decision Support System

International Command & Control Research & Technology Symposium

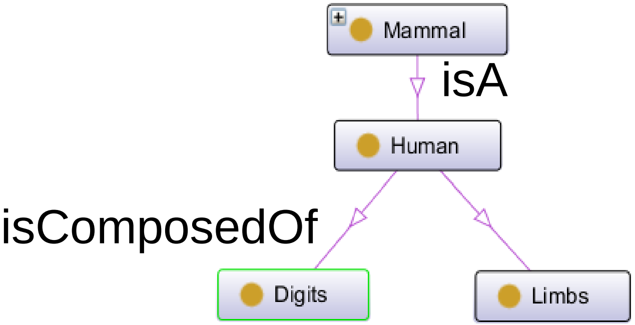
June 16-19, 2014

Richard J. Haberlin Jr. (EMSolutions)
Paulo C.G. da Costa (George Mason University)
Kathryn B. Laskey (George Mason University)

Background

Why Probabilistic Ontologies?

- Suppose an ontology of organisms contains the following classes and relationships:



Mammal

isA

Human

isComposedOf

Digits

Limbs

- Humans *usually* have:
  - 2 arms & 2 legs
  - 10 fingers & 10 toes
- However, if a man loses a limb….
  - Is he no longer human?

Premise of an argument can be uncertain (e.g. Humans have 2 legs): (in)validity of the argument imposes no condition on the certainty of the conclusion (an amputee is Human).

Probabilistic Ontology Defined

A *probabilistic ontology* is an explicit, formal representation of knowledge about a domain of application. This includes

Ontology

- Types of entities that exist in the domain;
- Properties of those entities;
- Relationships among entities;
- Processes and events that happen with those entities;

Uncertainty

- **Statistical regularities that characterize the domain;**
- **Inconclusive, ambiguous, incomplete, unreliable, and dissonant knowledge related to entities of the domain;**
- **Uncertainty about all the above forms of knowledge;**

where the term entity refers to any concept that can be described and reasoned about within the domain of application [Costa, 2005].

An ontology is an explicit specification of a conceptualization [Gruber, 95].

A probabilistic ontology extends a traditional ontology to represent uncertainty.

Our Solution

- A systematic approach to probabilistic ontology development
  - Facilitated through a reference architecture
    - Formalizes the application of the methodology
    - Extensible to various domains

- Reference Architecture for Probabilistic Ontology Development (RAPOD)
  - A generalized reference architecture designed to collect, catalogue, and define the components required for development of probabilistic ontologies and establish the criteria to be satisfied by any set of selected tools and methods

RAPOD provides a flexible solution

Reference Architecture

- Provides a blueprint for architects to develop specific solution architectures within a defined domain.
  - Template for development
  - Defines integral components and their relationships
  - Reduces development time and project risk
- Standardizes language among participants
- Provides consistency of development within a domain
- Provides a reference for evaluation
- Establishes specifications and patterns

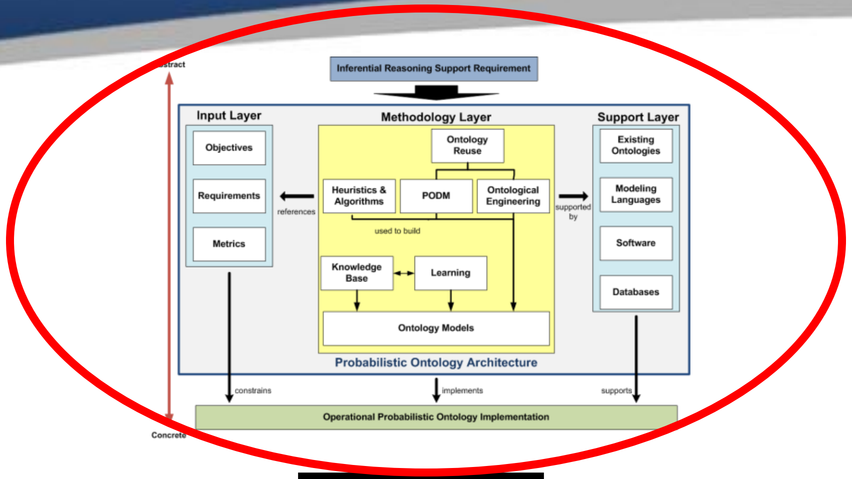
[A Reference Architecture is] "… an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions [OASD/NII, 2010]*."*

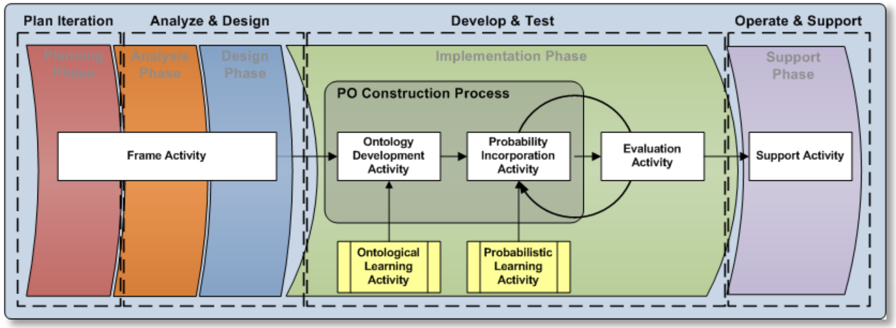
RAPOD Summarized

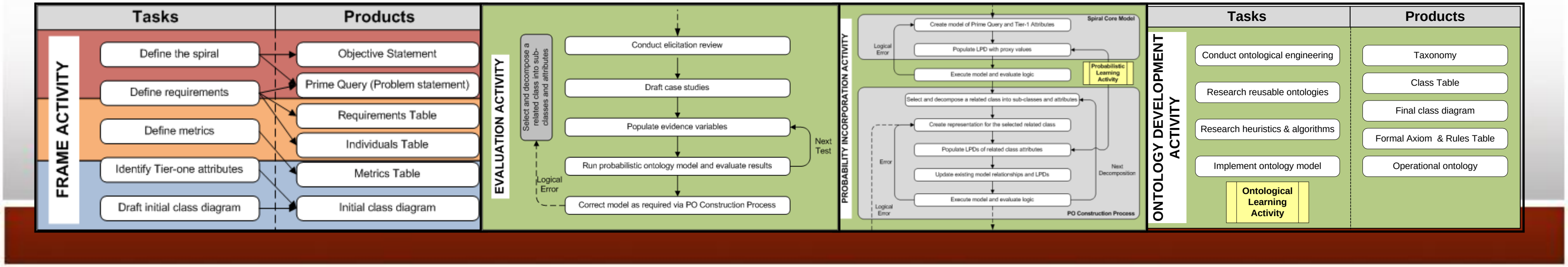
- Provides synergy of effort within the ST community
  - Identifies concepts, processes, languages, theories and tools
  - Synergizes effort of probabilists, logicians, decision analysts, computer scientists
- Spans knowledge, processes, models and tools necessary to engineer POs at a high level of abstraction
- Output defines a domain specific architecture that may be used to produce probabilistic ontologies in similar domain contexts

RAPOD output is an architecture

The RAPOD in Probabilistic Ontology Development







Tasks	Products
Define the spiral	Objective Statement
Define requirements	Prime Query (Problem statement)
Define metrics	Requirements Table
Identify Tier-one attributes	Individuals Table
Draft initial class diagram	Metrics Table
	Initial class diagram

EVALUATION ACTIVITY
Select and decompose a related class into sub-classes and attributes
Conduct elicitation review
Draft case studies
Populate evidence variables
Run probabilistic ontology model and evaluate results
Correct model as required via PO Construction Process
Next Test
Logical Error
PROBABILITY INCORPORATION ACTIVITY
Spiral Core Model
Create model of Prime Query and Tier-1 Attributes
Populate LPD with proxy values
Execute model and evaluate logic
Logical Error
Probabilistic Learning Activity
Select and decompose a related class into sub-classes and attributes
Create representation for the selected related class
Populate LPDs of related class attributes
Update existing model relationships and LPDs
Execute model and evaluate logic
Error
Next Decomposition
Logical Error
PO Construction Process
ONTOLOGY DEVELOPMENT ACTIVITY

Tasks	Products
Conduct ontological engineering	Taxonomy
Research reusable ontologies	Class Table
Research heuristics & algorithms	Final class diagram
Implement ontology model	Formal Axiom & Rules Table
Ontological Learning Activity	Operational ontology

Architecture for the Terrorist Identification Probabilistic Ontology (TIDPO)

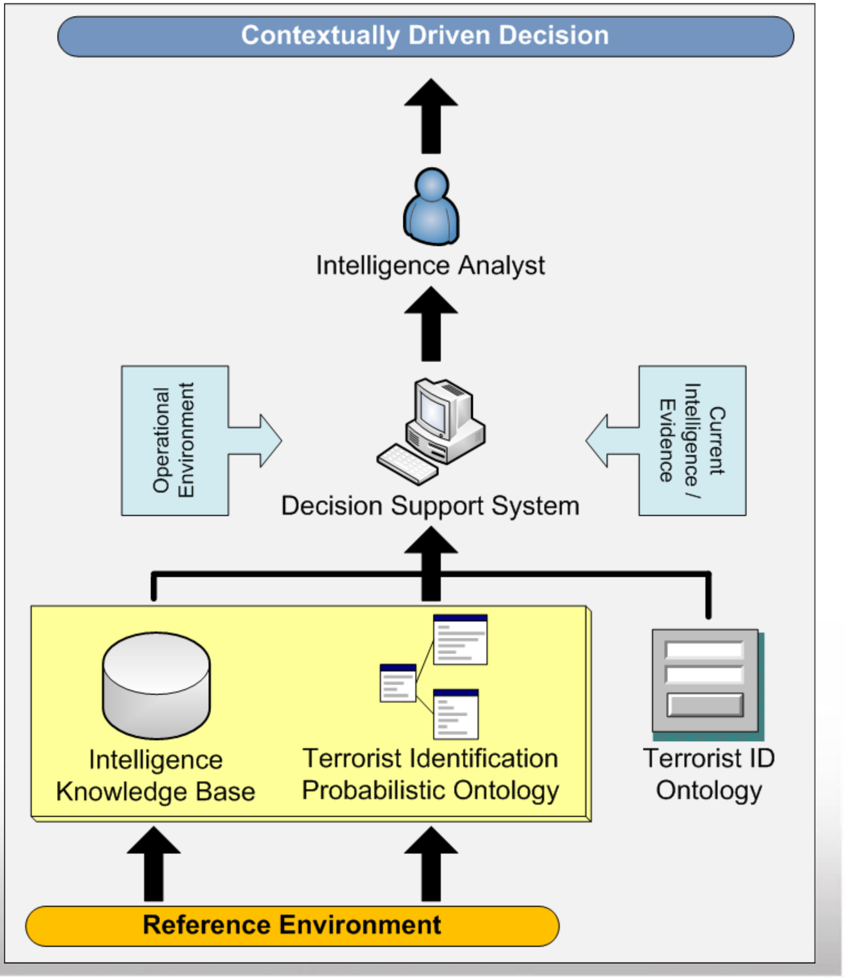
Terrorist Identification PO Concept

Objective

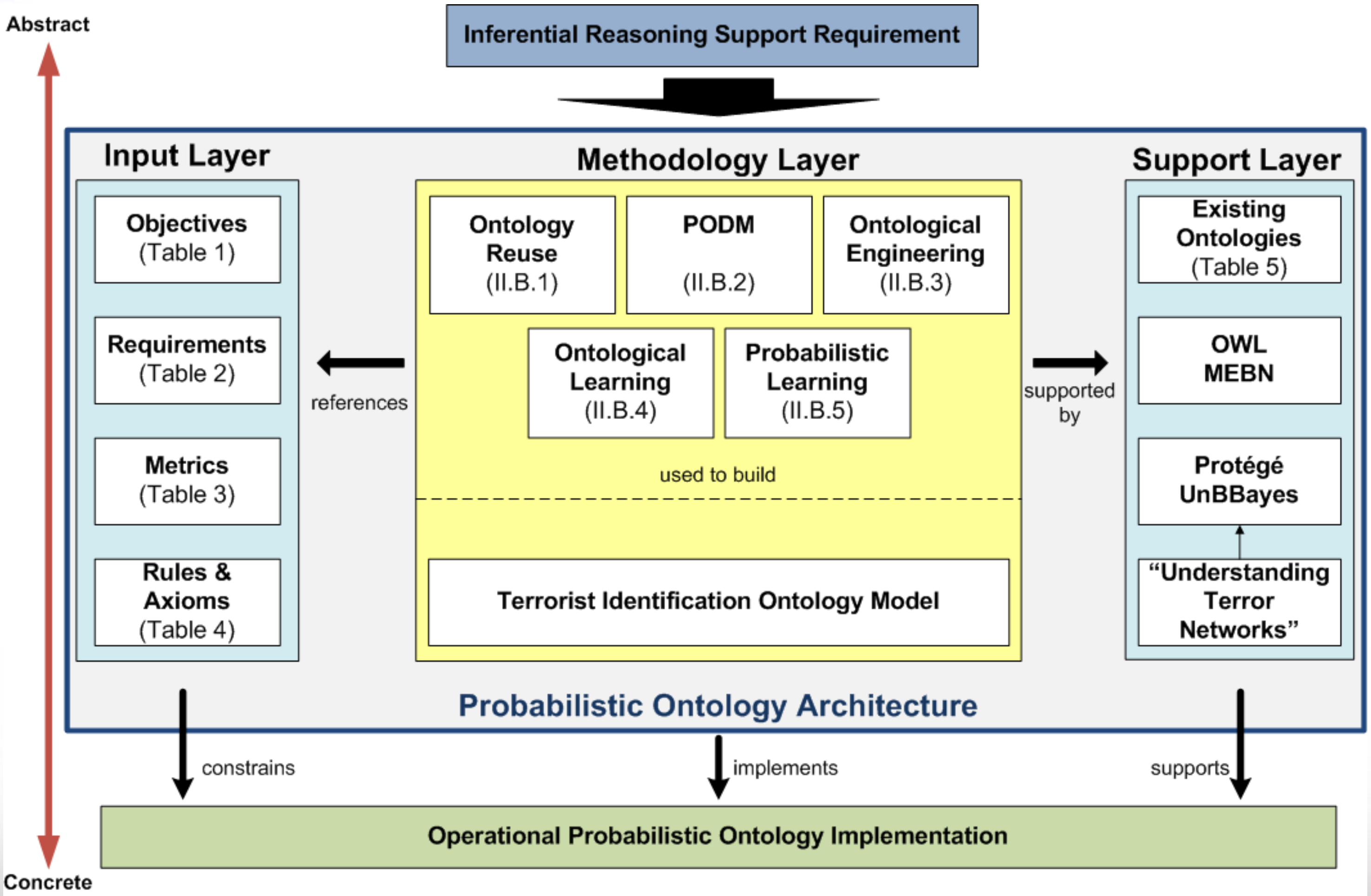
Develop a DSS that assists in determining if an individual is associated with terrorism.

“A DSS is a system under control of one or more decision makers that assists in the activity of decision making by providing an organized set of tools intended to impose structure on portions of the decision-making situation and to improve the ultimate effectiveness of the decision outcome [Marakas, 2003]*.”*

Concept Diagram


Contextually Driven Decision
Intelligence Analyst
Operational Environment
Current Intelligence / Evidence
Decision Support System
Intelligence Knowledge Base
Terrorist Identification Probabilistic Ontology
Terrorist ID Ontology
Reference Environment

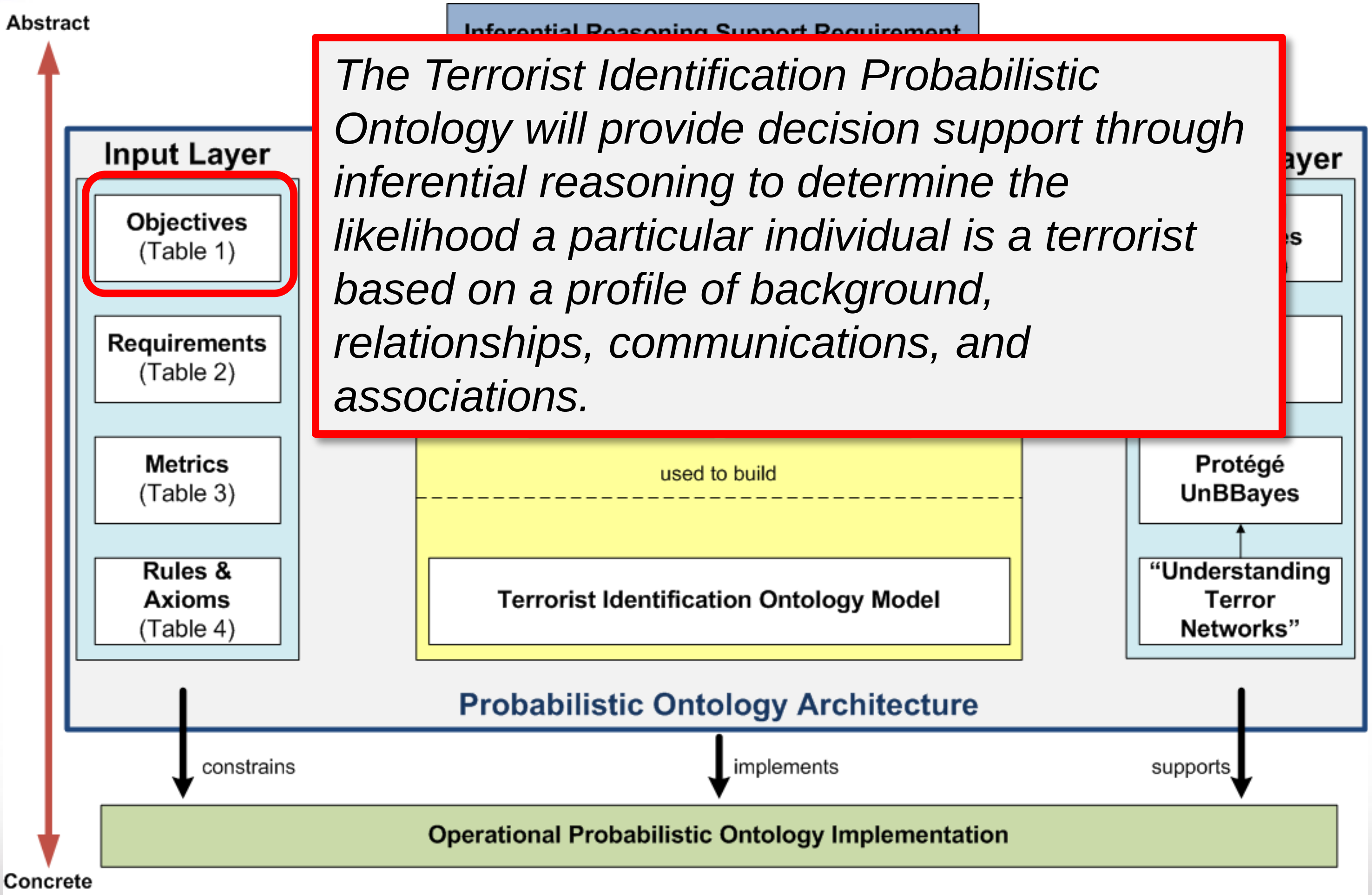
Architecture for Terrorist Identification Probabilistic Ontology



Abstract
Inferential Reasoning Support Requirement
Input Layer
Objectives (Table 1)
Requirements (Table 2)
Metrics (Table 3)
Rules & Axioms (Table 4)
references
Methodology Layer
Ontology Reuse (II.B.1)
PODM (II.B.2)
Ontological Engineering (II.B.3)
Ontological Learning (II.B.4)
Probabilistic Learning (II.B.5)
used to build
Terrorist Identification Ontology Model
supported by
Support Layer
Existing Ontologies (Table 5)
OWL MEBN
Protégé UnBBayes
"Understanding Terror Networks"
Probabilistic Ontology Architecture
constrains
implements
supports
Operational Probabilistic Ontology Implementation
Concrete

EMSolutions

Architecture for Terrorist Identification Probabilistic Ontology



Abstract

Inferential Reasoning Support Requirement

Input Layer

Objectives (Table 1)

Requirements (Table 2)

Metrics (Table 3)

Rules & Axioms (Table 4)

used to build

Terrorist Identification Ontology Model

ayer

s

Protégé UnBBayes

"Understanding Terror Networks"

Probabilistic Ontology Architecture

constrains

implements

supports

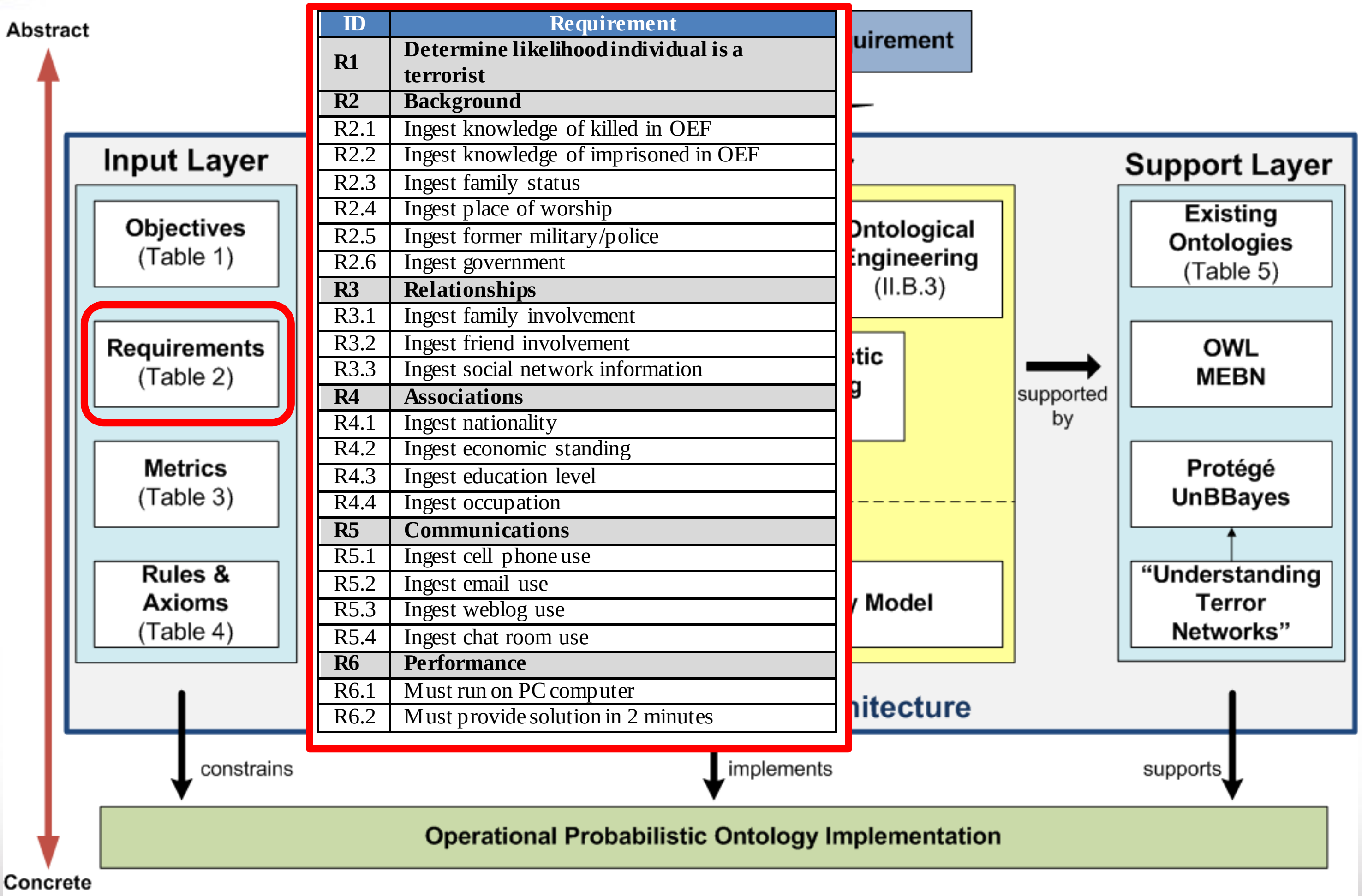
Operational Probabilistic Ontology Implementation

Concrete

The Terrorist Identification Probabilistic Ontology will provide decision support through inferential reasoning to determine the likelihood a particular individual is a terrorist based on a profile of background, relationships, communications, and associations.

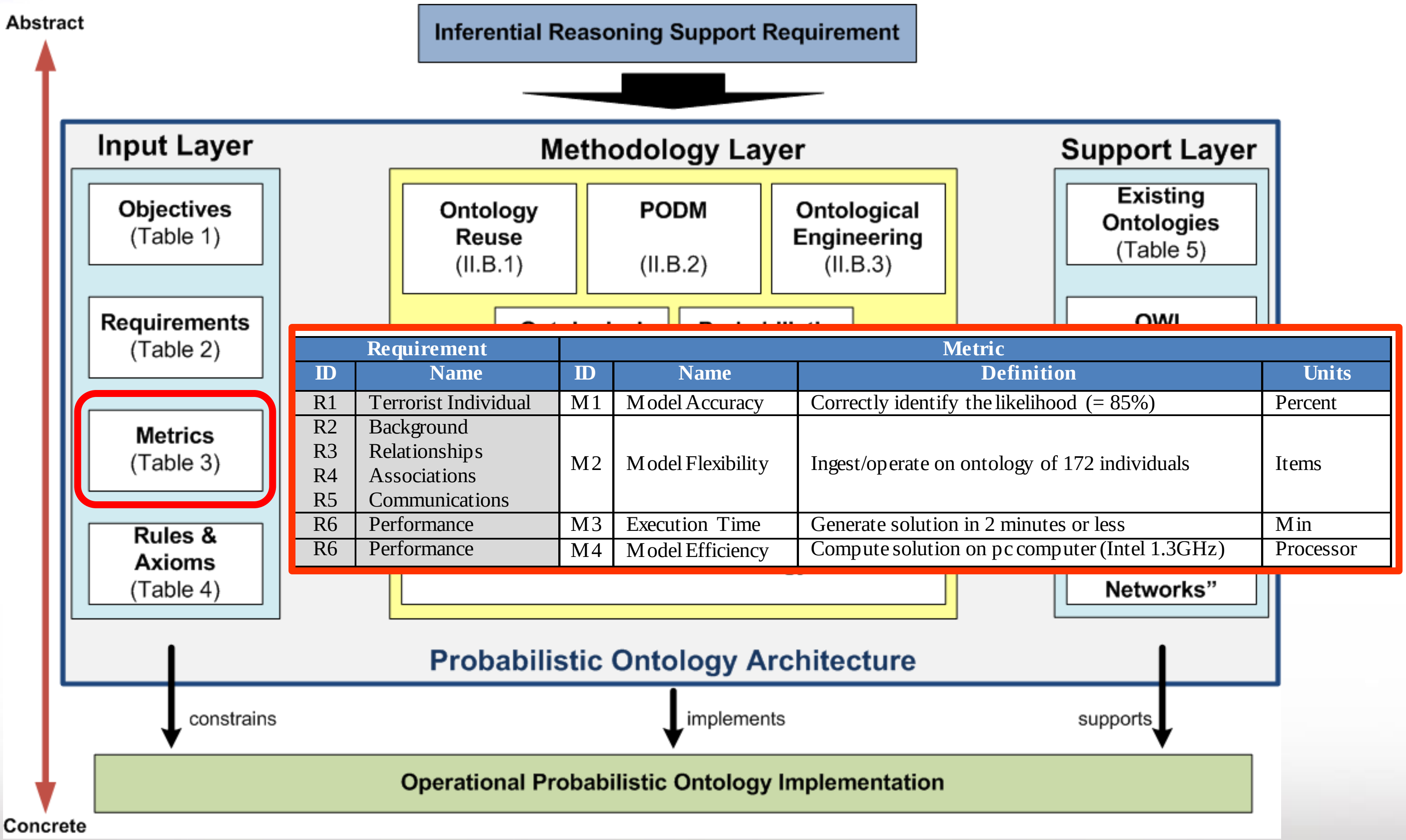
EMSolutions

Architecture for Terrorist Identification Probabilistic Ontology



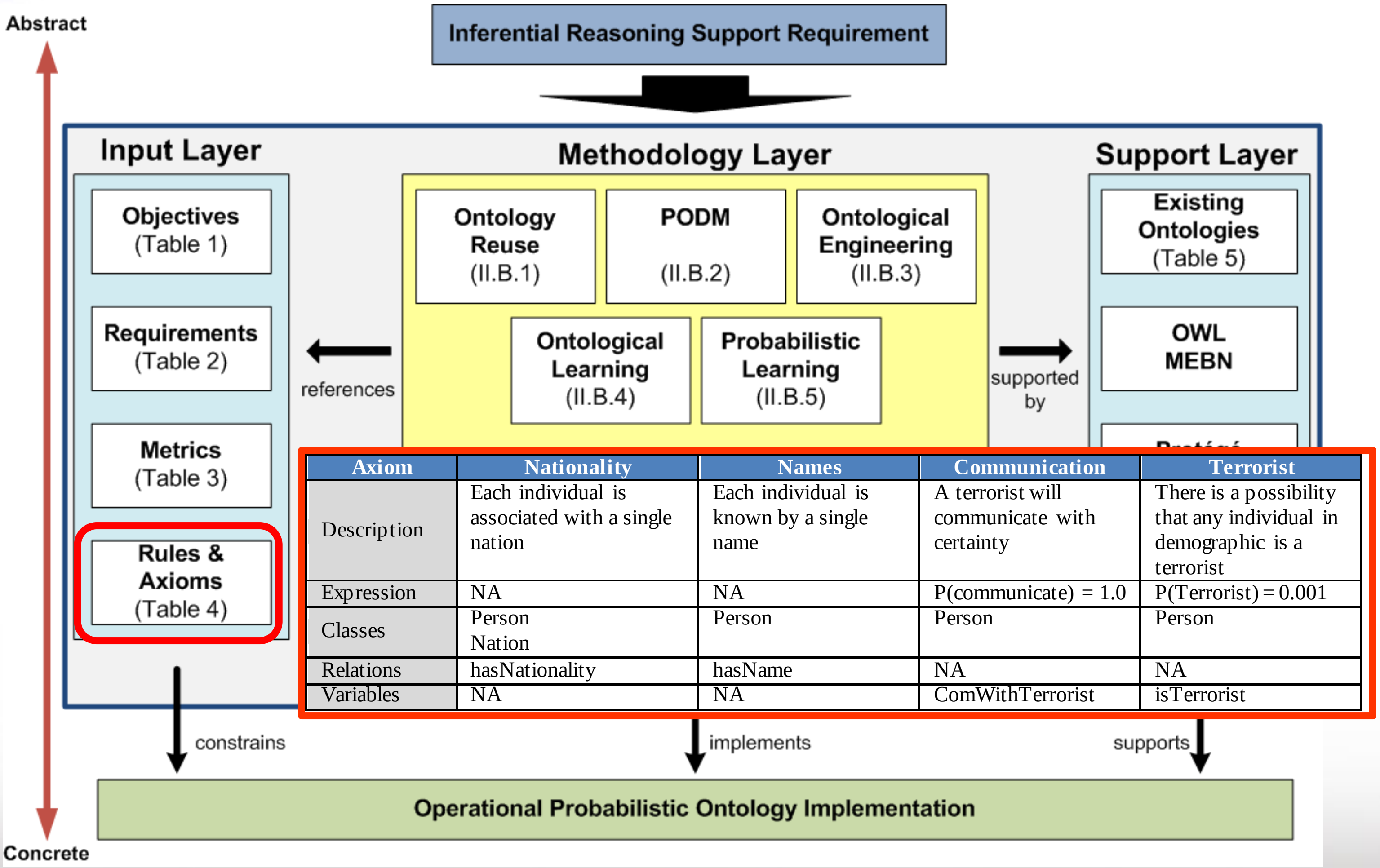
ID	Requirement
R1	**Determine likelihood individual is a terrorist**
R2	**Background**
R2.1	Ingest knowledge of killed in OEF
R2.2	Ingest knowledge of imprisoned in OEF
R2.3	Ingest family status
R2.4	Ingest place of worship
R2.5	Ingest former military/police
R2.6	Ingest government
R3	**Relationships**
R3.1	Ingest family involvement
R3.2	Ingest friend involvement
R3.3	Ingest social network information
R4	**Associations**
R4.1	Ingest nationality
R4.2	Ingest economic standing
R4.3	Ingest education level
R4.4	Ingest occupation
R5	**Communications**
R5.1	Ingest cell phone use
R5.2	Ingest email use
R5.3	Ingest weblog use
R5.4	Ingest chat room use
R6	**Performance**
R6.1	Must run on PC computer
R6.2	Must provide solution in 2 minutes

Architecture for Terrorist Identification Probabilistic Ontology



Requirement		Metric			
ID	Name	ID	Name	Definition	Units
R1	Terrorist Individual	M1	Model Accuracy	Correctly identify the likelihood (= 85%)	Percent
R2 R3 R4 R5	Background Relationships Associations Communications	M2	Model Flexibility	Ingest/operate on ontology of 172 individuals	Items
R6	Performance	M3	Execution Time	Generate solution in 2 minutes or less	Min
R6	Performance	M4	Model Efficiency	Compute solution on pc computer (Intel 1.3GHz)	Processor

Architecture for Terrorist Identification Probabilistic Ontology



Abstract

Inferential Reasoning Support Requirement

Input Layer

Objectives (Table 1)

Requirements (Table 2)

Metrics (Table 3)

Rules & Axioms (Table 4)

references

Methodology Layer

Ontology Reuse (II.B.1)

PODM (II.B.2)

Ontological Engineering (II.B.3)

Ontological Learning (II.B.4)

Probabilistic Learning (II.B.5)

supported by

Support Layer

Existing Ontologies (Table 5)

OWL MEBN

Protégé

constrains

implements

supports

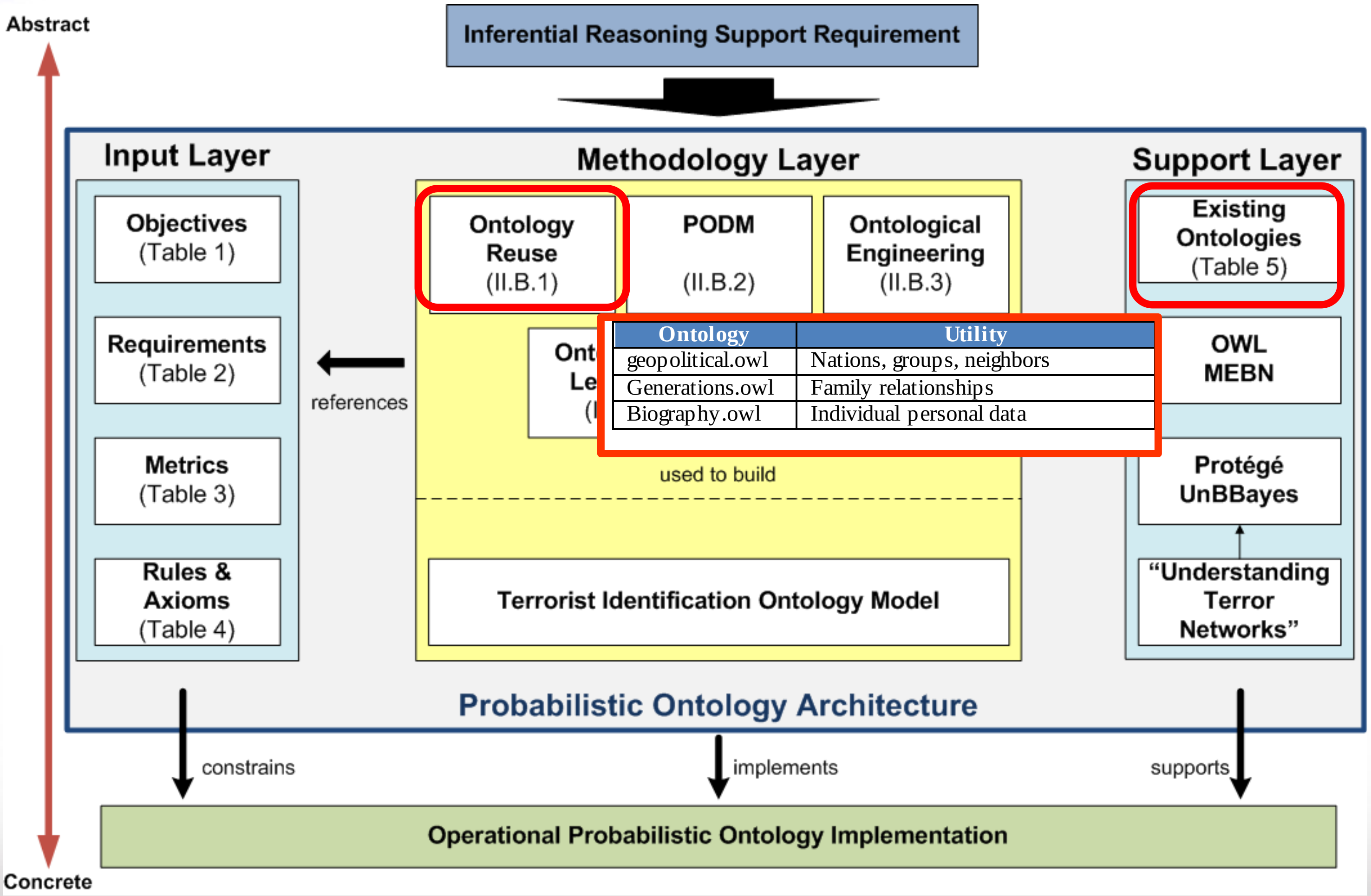
Operational Probabilistic Ontology Implementation

Concrete

Axiom	Nationality	Names	Communication	Terrorist
Description	Each individual is associated with a single nation	Each individual is known by a single name	A terrorist will communicate with certainty	There is a possibility that any individual in demographic is a terrorist
Expression	NA	NA	P(communicate) = 1.0	P(Terrorist) = 0.001
Classes	Person Nation	Person	Person	Person
Relations	hasNationality	hasName	NA	NA
Variables	NA	NA	ComWithTerrorist	isTerrorist

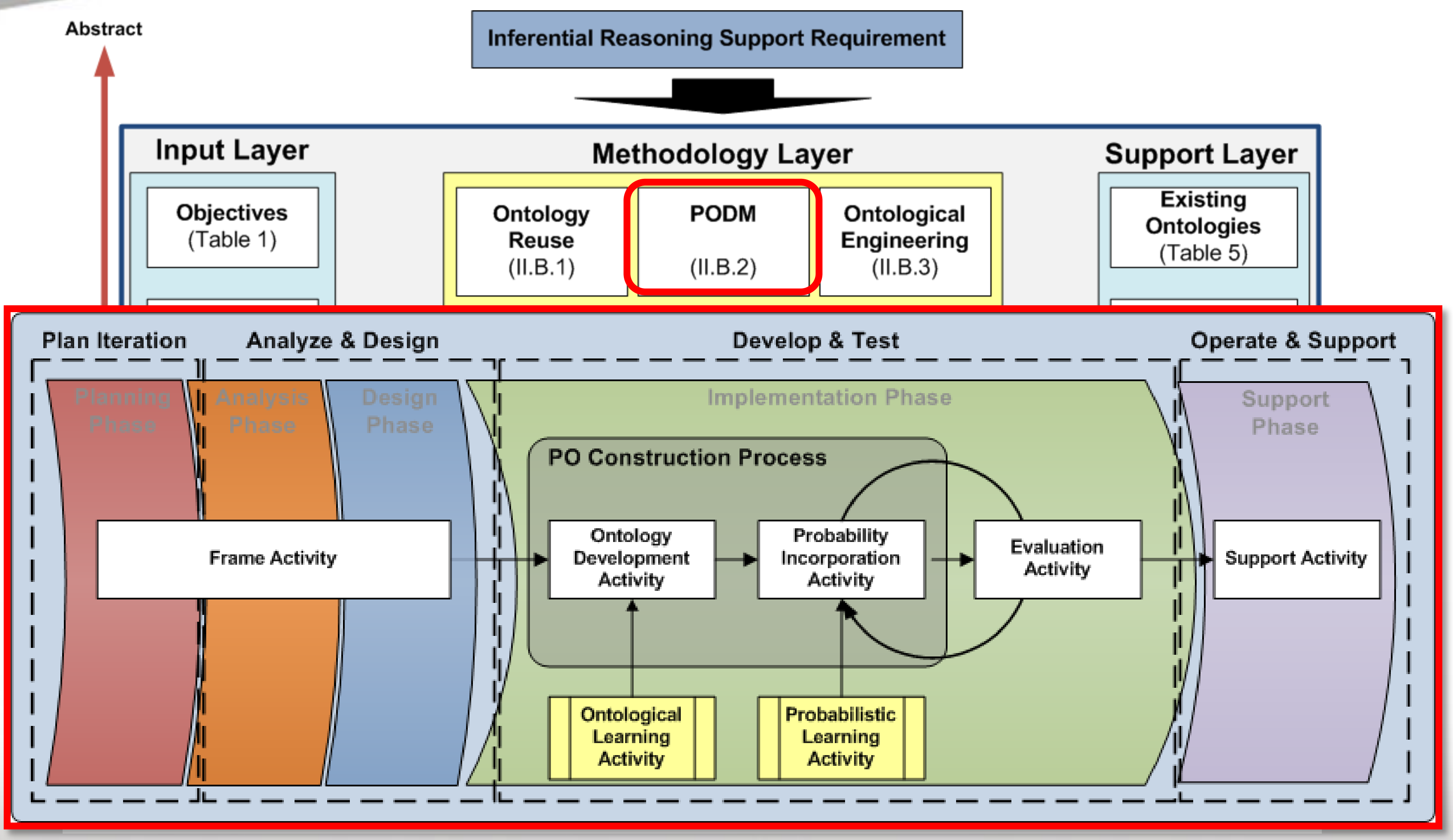
EMSolutions

Architecture for Terrorist Identification Probabilistic Ontology



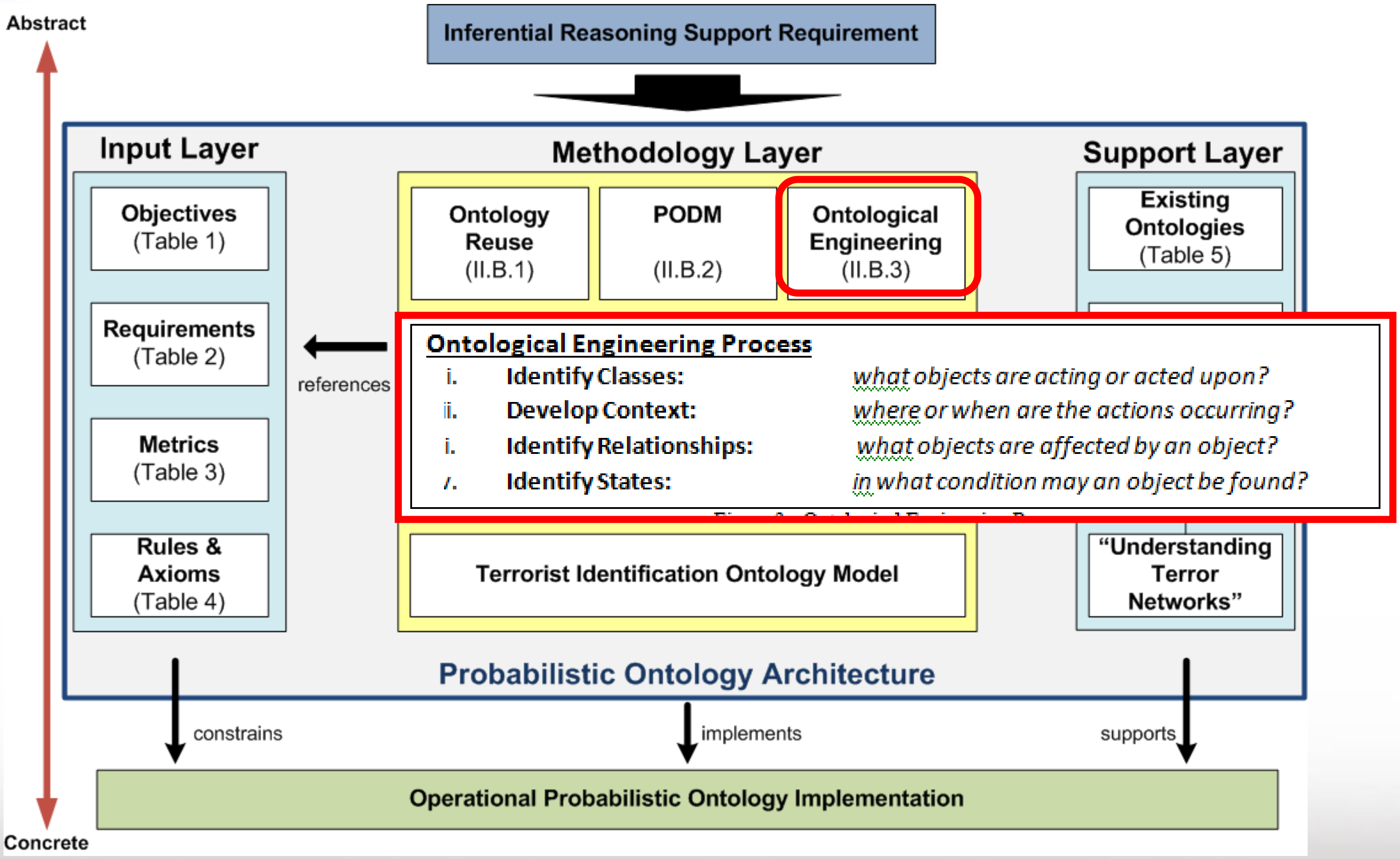
Ontology	Utility
geopolitical.owl	Nations, groups, neighbors
Generations.owl	Family relationships
Biography.owl	Individual personal data

Architecture for Terrorist Identification Probabilistic Ontology



EMSolutions

Architecture for Terrorist Identification Probabilistic Ontology



Abstract

Inferential Reasoning Support Requirement

Input Layer

Objectives (Table 1)

Requirements (Table 2)

Metrics (Table 3)

Rules & Axioms (Table 4)

references

Methodology Layer

Ontology Reuse (II.B.1)

PODM (II.B.2)

Ontological Engineering (II.B.3)

Ontological Engineering Process

i. **Identify Classes:** *what objects are acting or acted upon?*

ii. **Develop Context:** *where or when are the actions occurring?*

i. **Identify Relationships:** *what objects are affected by an object?*

/. **Identify States:** *in what condition may an object be found?*

Terrorist Identification Ontology Model

Support Layer

Existing Ontologies (Table 5)

“Understanding Terror Networks”

Probabilistic Ontology Architecture

constrains

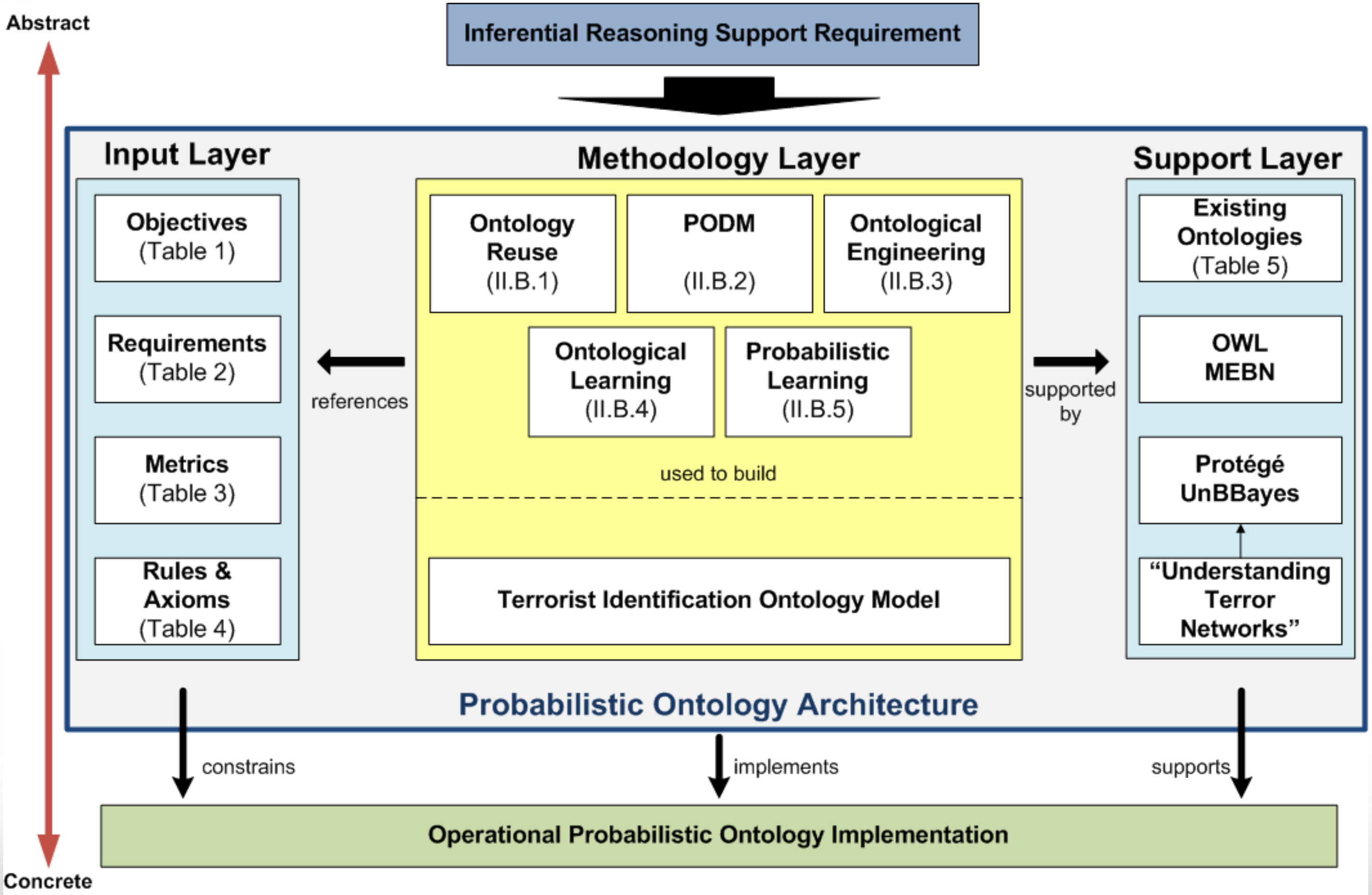
implements

supports

Operational Probabilistic Ontology Implementation

Concrete

Architecture for Terrorist Identification Probabilistic Ontology



Abstract
Inferential Reasoning Support Requirement
Input Layer
Objectives (Table 1)
Requirements (Table 2)
Metrics (Table 3)
Rules & Axioms (Table 4)
references
Methodology Layer
Ontology Reuse (II.B.1)
PODM (II.B.2)
Ontological Engineering (II.B.3)
Ontological Learning (II.B.4)
Probabilistic Learning (II.B.5)
used to build
Terrorist Identification Ontology Model
supported by
Support Layer
Existing Ontologies (Table 5)
OWL MEBN
Protégé UnBBayes
"Understanding Terror Networks"
Probabilistic Ontology Architecture
constrains
implements
supports
Operational Probabilistic Ontology Implementation
Concrete

EMSolutions

GEORGE MASON UNIVERSITY

Probabilistic Ontology Architecture for a Terrorist Identification Decision Support System

Richard Haberlin

Paulo Costa

Kathy Laskey

EMSolutions